

A FINAL PROJECT PERFORMANCE TEST FOR A SPANISH CONVERSATION CLASS AT A KOREAN UNIVERSITY

MARIA TERESA MARTINEZ-GARCIA

HANKUK UNIVERSITY OF FOREIGN STUDIES

Introduction

My name is Maria Teresa Martinez-Garcia and I am an assistant professor of Spanish at Hankuk University of Foreign Studies (Seoul, South Korea), where I teach Spanish classes. The classes range from communicative/elementary Spanish to advanced debate or essay writing. My background is in linguistics (Ph.D., University of Kansas), and my dissertation research (Martinez-Garcia, 2016) took a psycholinguistic approach to understanding bilingual activation, by exploring how differences in stress placement between English-Spanish identical cognates affect how adult learners of Spanish use stress as a cue for word recognition. My research interests include bilingualism, second language acquisition, speech perception and production, and pedagogical approaches to teaching pronunciation in the foreign language classroom.

The test I present here was used with my intermediate conversation class. Hankuk University of Foreign Studies (HUFSS) is a private research university, which specializes in foreign language education, offering 53 different language courses. The Spanish department is one of the largest departments at the university (together with the English and Chinese departments), with about 500 students enrolled seeking to graduate with a major or a minor in Spanish. Based on department policy, the intermediate conversation class has been created to help students reach a B1 level in Spanish speaking, following the standards of the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001). The class consists of 20 students who weekly meet for 2 hours, for a total of 16 weeks. When students take this class, they have normally taken two beginner conversation classes and most of the grammar and composition

classes. The latter amounts to four semesters of grammar courses and two semesters of composition. Apart from these classes, which are mostly taught in Spanish, these students have already taken most of the core courses in the Spanish department (e.g., History of Spanish Culture or Latin American Literature), which are normally taught in Korean, most learners' L1. In order to graduate with a major or a minor in Spanish, students need to complete six semesters of conversation in Spanish, which is the focus of the test presented here.

One of the challenges of teaching in South Korea is to get students to participate in conversation classes. My feeling is their grammar is good, and that they can write complex essays about almost any topic, even from their first semester of formally learning Spanish. However, given the East Asian culture in which they have grown up, they are mostly used to classes in which the teacher/professor lectures and they only need to take notes, so it is normally difficult to change the dynamics of the class so that they are the ones doing most of the talking. With this project, I wanted students to do some real talking (they had the opportunity to plan and prepare out of class), but still let them decide about what they want to talk, to increase the communicativeness of the project.

The Test

Learners receive the test, the teacher's grading rubric (see Table 4-30), and the student's grading rubric (see Table 4-31), in Spanish. The instructions, together with the rubrics, are handed to students in the middle of the semester, to make sure students enough time to complete the project. After discussing the details of the test in class, students form groups and start their independent work preparing the final project (meeting to discuss the topic to be covered, recording the video, and preparing their final presentation). During this time, students are encouraged to attend my office hours to get help on the different steps of the project. Office hour topics range from brainstorming possible ideas of topics to drafting their final presentation.

When scoring the final projects (video and oral presentation) it is really important for me to see whether students would meet the requirements to pass the DELE B1 oral exam (see Baztán, Torrecillas, Cuadrado, Guerrero, & Molero, 2015). For that reason, and knowing how familiar Korean students are with memorizing information, I include an extra question for which they could not prepare and that would give me more information regarding their real level of proficiency in the language in terms of understanding and speaking. Moreover, I evaluate their complete performance using the grading scales created by the Instituto Cervantes to

evaluate the speaking section of the B1 DELE exams, and which focus on coherence, fluency, linguistic scope (i.e., lexis), and correctness (see Instituto Cervantes, 2019).

PROYECTO FINAL

El Proyecto Final se debe completar en parejas o grupos de tres personas (máximo). Para completar el proyecto final, debéis responder a la siguiente pregunta:

¿Qué significa esta universidad para ti?

El proyecto final evaluará tanto la gramática/vocabulario como la presentación y la originalidad, y consiste de dos partes, relacionadas entre sí.

1. PARTE 1 (vídeo: 40 puntos)

Instrucciones: Vais a responder a la pregunta principal: ¿Qué significa esta universidad para ti? grabando un vídeo en el que mostráis porqué la universidad es importante para vosotros.

- Detalles del vídeo: El vídeo debe mostrar imágenes que representen vuestra respuesta a la pregunta. No debe contener voz (está bien incluir algunas palabras (en español) o sonidos de voces de fondo). Podéis incluir música si queréis (los ejemplos os darán una idea).
 - Durante la **presentación** en clase es cuando incluiréis la “voz” y las palabras para responder a la pregunta y explicar las imágenes.
- Duración del vídeo:
 - Grupo de 2 personas: Vídeo de unos 4 minutos.
 - Grupo de 3 personas: Vídeo de unos 6 minutos.

Ideas: A la hora de grabar el vídeo pensad en temas como: la enseñanza, el crecimiento personal, la integración, nuevas oportunidades, las relaciones personales, la amistad, el futuro (laboral y personal), etc.

Ejemplos de vídeos:

- ¿Qué significa la verdadera amistad para ti?
<https://www.youtube.com/watch?v=ozTrgZdntwI>
- ¿Qué significa la educación para ti?
<https://www.youtube.com/watch?v=Un5msddQl6U>

2. PARTE 2 (presentación: 60 puntos (50 presentación + 10 comentarios de compañeros))

Instrucciones: En grupos, vais a presentar el vídeo. No se puede leer (aunque podéis tener notas con vosotros para ayudaros) y todos los miembros del grupo deben hablar más o menos durante el mismo tiempo.

Formato:

1. **Presentación** (unos 2 minutos): Presentación de los integrantes del grupo y de la motivación del vídeo (sobre qué temas trata y por qué pensasteis que eran temas importantes).
2. **Vídeo** (duración del vídeo): En este punto, debéis poner la voz en *off* al vídeo. Como narradores de la historia, debéis explicar qué está pasando en el vídeo.
 - a. No es necesario hablar durante los 4 (o 6) minutos del vídeo, pero sí la mayor parte del mismo.
3. **Conclusión** (unos 2 minutos): Resumen del proyecto, incluyendo cómo ha cambiado vuestra perspectiva de la universidad (o no) desde que sois estudiantes.
4. **Pregunta** (no preparada): Pregunta no preparada, pero relacionada con el proyecto presentado. Todos los miembros del grupo deben responderla individualmente. Ejemplos de preguntas:
 - a. ¿Por qué decidiste estudiar en esta universidad?
 - b. ¿Volverías a tomar la decisión de estudiar en esta universidad?

- c. ¿Qué cambiarías del tiempo que has pasado en esta universidad?

Table 4-30: The teacher's grading rubric Spanish version

RÚBRICA DE LA PRESENTACIÓN (profesora)

	Supera las expectativas / A	Cumple las expectativas / B	En progreso / C	No cumple las expectativas / D-F
Léxico	Utiliza de forma apropiada y con efectividad el vocabulario necesario.	Utiliza de forma apropiada y bastante efectiva el vocabulario necesario.	Utiliza el vocabulario apropiado la mayor parte del tiempo; palabras en otros idiomas.	No utiliza el vocabulario de forma adecuada; el uso de otros idiomas afecta la comunicación.
Gramática	La gramática básica está toda perfecta.	La gramática básica está casi perfecta.	La gramática básica presenta errores importantes; influencia de otras lenguas.	La gramática presenta muchos errores; influencia de otras lenguas.
Fluidez	No existen problemas de pronunciación importantes. No hay demasiadas pausas y el discurso es fluido y coherente.	Hay pocos problemas de pronunciación (4). Hay pausas y muestra dudas, por lo que el discurso no es fluido.	Hay bastantes problemas de pronunciación (+4). Hay pausas y muestra dudas, por lo que el discurso no es fluido.	Los problemas de pronunciación afectan la comprensión. Existen pausas largas, dudas, y el discurso no es fluido.

Comprensión y producción	Completa la tarea con éxito. La gramática no se limita a estructuras básicas. Hay comunicación e interacción entre ambas partes.	Buen entendimiento de la tarea y utilización del lenguaje apropiado. Hay comunicación e interacción entre ambas partes.	Entiende la temática del proyecto, aunque la tarea no está completa o no es lógica. La participación entre los miembros no está compensada.	El grupo no entiendo el proyecto y la producción está incompleta o no tiene lógica. La información proporcionada es demasiado básica o incluso inexistente.
Originalidad	El proyecto es único, no se parece a los demás. Muestra creatividad, es único y fresco.	El proyecto funciona, pero no es único ni original. Tiene bastantes componentes repetitivos.	El proyecto no está bien organizado. Tiene muchos componentes repetitivos.	Los estudiantes apenas prepararon el proyecto final.

Most universities in South Korea follow a relative grading system, which is a government-imposed grading curve. While it depends on schools, this grading curve demands that at least 30 percent of students are destined to receive a grade of C+ or lower. So, even when students deserve to get a higher level, the grading system at the university does not allow professors to give more than a certain percentage of As, of Bs, and of Cs. This means that the difference between a student getting an A and a student getting a C may be a couple of points (depending on how good the class is, someone with 91/100 points at the end of the semester may end up getting a C).

Thus, in Korea, getting a C is almost as bad as failing the class. Students often request to know in advance how exactly each individual project is going to be evaluated. Having these rubrics at hand help them identify in which aspects they may be weaker and how they may be able to improve their performance on the project and, hopefully, their grade.

Table 4-31: *The students' grading rubric Spanish version*

RÚBRICA PARA EVALUAR LAS PRESENTACIONES
(estudiantes)

Tu nombre: _____ Presentación número: _____

<p>1. Del 1 al 10, siendo 10 la máxima nota, ¿cómo evaluarías la originalidad del vídeo de este grupo?</p> <p style="text-align: center;">1 2 3 4 5 6 7 8 9 10</p>
<p>2. Del 1 al 10, siendo 10 la máxima nota, ¿cómo evaluarías la claridad de la presentación de este grupo?</p> <p style="text-align: center;">1 2 3 4 5 6 7 8 9 10</p>
<p>3. Del 1 al 10, siendo 10 la máxima nota, ¿cómo evaluarías la preparación de este grupo?</p> <p style="text-align: center;">1 2 3 4 5 6 7 8 9 10</p>
<p>¿Tienes algún comentario más para el grupo?</p>

English Translation of the Test and Scoring Rubrics

Below please find the English translations of the test sheet, the teacher's grading rubric (Table 4-32) and the students' grading rubric (Table 4-33).

Table 4-32: The teacher’s grading rubric English version

RUBRIC TO EVALUATE PRESENTATIONS (teacher)

	Above expectations / A	Meet the expectations / B	In progress / C	Does not meet the expectations / D-F
Vocabulary	Employs appropriately and effectively the required vocabulary.	Employs appropriately and quite effectively the required vocabulary.	Employs the appropriate vocabulary most of the time: words in other languages.	Doesn't use the vocabulary in a proper way; overuse of foreign languages, which affects communication .
Grammar	Basic grammar is perfectly used.	Basic grammar is almost perfectly used.	Basic grammar present important errors; influence from other languages.	Grammar presents a lot of mistakes; influence from other languages.
Fluidity	There are no important problems in pronunciation. Barely any pause and fluid, coherent discourse.	There are some errors in pronunciation (4). There are some pauses, and shows doubts, so the discourse is not fluid.	There are a lot of pronunciation problems (+4). A lot of pauses, which show doubt, so the discourse is not fluid.	The pronunciation problems affect comprehension . Long pauses, doubts, and the discourse is not fluid.
Comprehension and production	Successful completion of the task. The grammar is not limited to basic structures. There is communication between the members of the group.	Good understanding of the purpose of the task, and appropriately use of the vocabulary/grammar . There is communication between the members of the group and the audience.	Understands the purpose of the task, although the project is not complete or logic. The members of the group do not participate equally in the completion of the task.	The group does not show understanding of the purpose of the project and the work is not complete or logic. The information presented is too basic or inexistent.

Originality	Unique project, which stands out from the rest. It shows creativity, freshness and uniqueness.	The project meets the requirements, but it's not unique or original. It has components that are repeated in all the other projects.	The project is not well-organized. It is very repetitive.	Very little preparation shown by the students.
-------------	--	---	---	--

Table 4-33: The students' grading rubric English version

RUBRIC TO EVALUATE PRESENTATIONS (students)

Name: _____
number: _____

Presentation

<p>1. Between 1 and 10, being 10 the highest score, how would you evaluate the originality of this group's video?</p> <p style="text-align: center;">1 2 3 4 5 6 7 8 9 10</p>
<p>2. Between 1 and 10, being 10 the highest score, how would you evaluate the clarity of this group's presentation?</p> <p style="text-align: center;">1 2 3 4 5 6 7 8 9 10</p>
<p>3. Between 1 and 10, being 10 the highest score, how would you evaluate this group's preparation?</p> <p style="text-align: center;">1 2 3 4 5 6 7 8 9 10</p>
<p>Do you have any feedback for the group?</p>

Contributor’s Questionnaire Responses

Section One: Test Planning and Writing

Why did you write the test? What were the purposes of the test?

This test was created to evaluate students’ progress in a conversation class. My students are used to either memorizing what they have to say, or to write a lot in their exams. However, no other professor had tried to measure their real speaking performance, and I tried to change that with my final project.

The evaluation of the course was divided between a midterm oral exam (a personal interview with me to help me establish a “starting point” from which to evaluate students’ progress), attendance and participation, weekly homework (including writing, reading, listening and speaking activities), and this final project. Although students’ final evaluation followed those four aspects, this final project accounted for 40% of their final grade.

The main idea for me was to create a test that would allow me to have a real assessment of their progress in their speaking skills. Most of the students had memorized parts of the midterm oral exam, and so I was not able to evaluate their real production. I used this test to measure their progress and their speaking skills, and each student received a final report including information regarding his/her individual progress, his/her participation in the group, and the feedback students gave to the group regarding how original the project was and how well they had prepared to complete this assignment. See Table 4-34.

Table 4-34: *The final project evaluation form*

COMMENTS ON YOUR FINAL PROJECT AND GRADE

Notes and commentaries:

1. Evaluation and average grade give to your group by your classmates (10% of final grade)	
• Between 1 and 10, being 10 the highest score, how would you evaluate the originality of this group’s video?	/ 10
• Between 1 and 10, being 10 the highest score, how would you evaluate the clarity of this group’s presentation?	/ 10
• Between 1 and 10, being 10 the highest score, how would you evaluate this group’s preparation?	/ 10

Feedback given to you by your classmates:	
AVERAGE GRADE:	/ 10
2. Evaluation and average grade given to you by the professor regarding your group work: It includes the originality of the video and the division of the work between the members of the group (40% of final grade)	
Feedback:	/ 10
3. Evaluation and average grade given to you by the professor regarding your individual work (50% of final grade)	
Feedback:	/ 10
FINAL GRADE:	/ 10

How did you decide how many subtests to write? Was there some correspondence between a subtest and a course objective, or was there some other reason to create the subtests you did?

In my final project, there is not really something like a “subtest.” There are two parts to this test. The first part is a video that students would need to record and prepare in pairs/small groups, and the second part represents their presentation of their video, putting in the voice “in-off” in front of everybody in the classroom and answering my final question (for which they could not prepare).

My decision on making this division was to give students time to prepare the video, working in small groups to really think about the topic and discuss among them the vocabulary they would need to use, the structures they could use, and the main ideas. Although not directly stated, this was my way to make sure they were preparing for the final presentation, without directly memorizing what they would need to say later on. The second part, which is also the one that obtains the highest credit, is the part in which they need to speak in public and that they cannot fully prepare (their personal answer to my final question).

Did you write as many items or test tasks as you needed, or did you write more than you needed, then pare the number down? How did you decide which items or test tasks to keep? How did you decide which items or test tasks to discard?

I honestly started by including three parts of the video, the first of which included a writing section. I was going to ask them to write down their answer to the specific question they had and for which they had to give an answer using the video/in-class presentation. However, after thinking about it, I realized that this task was not really related to the purpose of the class (improve their conversation skills) and so I decided to not include it. Rather, I encouraged students to really work on preparing the video for the presentations, and, indirectly, pushing them to prepare the vocabulary/grammar they would need for the presentation.

Did you have one version of your test, or did you create a second equivalent version?

I created several versions (with different topics, including the writing section versus deleting it, etc.), but I ended up just administering the one I am presenting together with this questionnaire. All students completed the same final project. In the future, I hope to be able to use the different versions originally prepared in future classes.

Were you concerned at how long the test would take to administer?

Not so much about how long it would take to administer, but how much time it would require for students to complete this project. The hardest part for them seemed to be to come up with a topic they would like to discuss and to prepare the video, but students said once they had a topic, they were done with it (they prepared the scripts and recorded the video) within a couple of days.

Were you concerned at how long the test would take to score?

Honestly, I did not think about this part when I was preparing the test, something that I regretted later on, when I started to evaluate the final projects. The quality of the videos was higher than I had anticipated, and students had worked really hard on their part of the presentation. In fact, it was the way in which they answered my final question that made the difference when evaluating the work of students.

Were you concerned how you might use the test items themselves for learner feedback?

From the beginning I knew I would be able to use my final question as a clear assessment of their conversation skills' progress. I would be able to determine whether they could understand my question (each group/student received a different question) and how they would be able to address it with respect to their level of proficiency and in relation to their "starting point" (the midterm oral exam).

Did you consider having your students take your test on a computer? Why or why not? If not, what were your concerns?

No. While there may be options to do it that way, I still prefer the in-class presentation method. I think the in-class presentation puts everybody at the same level in terms of external factors that could affect their performance. For example, some students feel more comfortable in front of a camera than others, thus those who are not so familiar/comfortable with talking in front of a computer could be at a disadvantage, not because they are less proficient in Spanish, but because they feel uncomfortable in that given situation. To avoid these external factors from introducing bias into my students' performances (and grades), I preferred asking everybody to do the same procedure.

Did you consider making the test an open book test?

With this type of test, having the book with them would not be particularly helpful, so I did not even consider the possibility of letting them bring their books.

Did you consider allowing learners to use additional sources such as dictionaries, or their notes, while taking the test?

I allowed them to bring notes with them. The notes helped them during the presentation, but they were asked not to read them, but just have them as support. The notes did not help during the final part of the presentation (my personal question).

Did you plan to allow learners to re-take a test for improvement? The same test, or a different test?

No. I gave them around two months to prepare for this final project, and they all knew that it was the most important part of their final evaluation and that they needed to prepare as carefully as possible.

What sources did you draw from for your test items?

In 2017, I participated in the Fundamentals of Project-Based Language Learning Online Institute, offered by the National Foreign Language Resource Center (NFLRC) at the University of Hawaii at Manoa (NFLRC, 2019). In this online institute, we learned about newer ways of approaching the teaching of foreign languages and content topics in a foreign language, in a way that is more meaningful for students. For example, we learned about how using topics interesting for students would make them feel more engaged with the learning process itself. Thus, the final product would be closer to students' actual level of proficiency because they did their best with something they like, rather than repeating what they had previously memorized.

Based on what I learned in the institute and some of the ideas that were discussed during the online lectures, I created this final project. The topic of education, for example, was discussed in class, but specifically asking them my personal question about what the university means to them was my idea. I thought about this specific question because it is a topic that is important for them right now. For instance, we do not have a library, as it is under renovation, and most of the buildings do not have an elevator and students with disabilities cannot attend those lectures, etc. I thought it would be interesting for them to look back at their time at the university and consider whether it was positive, negative, and what aspects they consider important.

When evaluating them I did try to follow some of the guidelines or descriptors proposed by ACTFL (American Council on the Teaching of Foreign Languages, 2012a). That is why I created a task that would engage them in free production (at least the last part of the presentation), while motivating them to carefully prepare for the presentation.

Which test item formats do you prefer to use?

I normally ask students to complete tests that include short answers, fill-in-the-blank, and matching tasks. However, none of these tasks made sense to evaluate students' conversation abilities. That is why I decided to change the format of the final project in this case. I would describe my final question as a short answer type of task, targeting free production.

What types of learner knowledge do you believe you are capturing in your test?

My aim was to try to capture the level of grammar and vocabulary that Spanish learners are expected to have at an intermediate level of proficiency. Again, my aim was to capture the knowledge they have and the use they can make of it in a context in which they cannot fully prepare their exam (like,

for example, memorizing sentences and just copying them when completing the exam).

What learner skills do you believe you are capturing in your test?

My aim is to capture learners' speaking and listening skills (primarily speaking skills). I think this test managed to capture this specific set of skills, mostly with the last question, which targets free comprehension and production.

If you wrote a performance test (where learners have to converse or present a topic, or write something above the sentence level), how did you get the ideas for what task learners had to do for the test?

I got my ideas from the real experiences that my students are having while being students at the university, and the ideas that were discussed in an online course, the Fundamentals of Project-Based Language Learning Online Institute, described above.

How did you get ideas on how to score learners' performances (the scoring criteria)?

This was one of the hardest parts, when I was creating this final project. I looked for ideas online, as well as ideas from the Online Institute mentioned earlier and the courses where I worked as a teaching assistant in the U.S.A. I wanted to create something that was unique and principled, but I also wanted to make sure that I was being fair and evaluating students' final project in a way that really captured the work done and the progress made during the semester.

If you used points on a scale for scoring learners' performances, how did you decide on how many points on a scale to use?

I only used a scale in the feedback that students provided to their classmates (Table 4-34). I just used a 1 to 10 scale, because I felt it was easier for students to relate this scale with the grades they normally obtain in class, and also because it was easier to calculate the final percentage obtained in this specific aspect of the evaluation of the final project.

Were you concerned about whether you could get a colleague to help you score learners' task performances?

Consider: Did you have ideas about how to ensure score consistency, such as having two scorers (you and a colleague), or scoring learners' performances by yourself on two different occasions?

Asking a colleague was not an option, as that is not the way in which things are done in South Korea. My colleagues experience a great amount of work towards the end of the semester, so I did not want to add them extra work by helping me evaluate my students. I knew I had to evaluate everything by myself, so I decided to evaluate the final projects myself on two different occasions.

On the first occasion, I gave them a score shortly after they had finished their presentation and I had been taking notes on their performance. Once I was back at home, I evaluated everybody's performance a second time, this time randomizing the order in which I reviewed the presentations/videos and I looked at my own notes.

Did you plan to give the scoring criteria to the learners for future self- or peer-assessment? Did you make another, perhaps simpler or shorter, version of the scoring criteria for learners to use?

From the very first moment, I gave them the scoring criteria in Spanish, including the students' evaluation form (Table 4-31), together with the instructions on how to complete the final project. I wanted to make sure they would understand how I was going to be evaluating their work even before they started working on it. Understanding how important the presentation was would make them more conscious regarding that part of their project and more motivated to do their best. To make sure they understood the instructions and rubrics, we went over them together during part of one of the class sessions.

Did you seek help from a peer to clarify what your test items or tasks were measuring?

Consider: Did you state at any point what you wanted to measure in your test?

Did you ask another teacher to compare your test items with what you said you wanted to measure?

Did you make any changes to your test or items as a result of your colleague's feedback?

I asked for feedback from former colleagues who have used tasks like the one I used. I wanted to make sure that 1.) My task was indeed appropriate for an intermediate level, 2.) The amount of work was fair as a final project targeting conversation skills, and 3.) That the wording used in the instructions was clear and concise. I adjusted the wording of my instructions, and I decided to remove the writing section of the project after hearing their feedback.

Did you compare your test to the lessons that learners had?

I did, to make sure I was not asking them to do something beyond their current level of proficiency.

Did you compare your test to the textbook or other materials learners used?

I did, to make sure I was not asking them to do something beyond their current level of proficiency.

Adapting existing tests

Are you required to use specific tests in your program?

No, I have complete freedom in the decision of what type of tests to use. However, I am expected to use a test that targets the main content covered in the class (e.g., speaking test for the conversation class) and keep some sort of record in case the university wants me to explain my decision on certain grades.

Did you inherit your test or parts of your test?

No, I created everything by myself, although it is true that I gathered ideas from the examples we were given at the National Foreign Language Resource Center Institute.

Section Two: Test Administering and Scoring

Were you concerned about test security? What did you do to ensure test security?

I was really concerned about test security. One of my problems is that I could not make sure that students were not copying their ideas from other sources, for example, by reading/watching what other students had done with similar projects. That is why I decided to include the last question, which I personalized for each student/group, based on what they said in their presentation. This was my way of trying to avoid any type of cheating, as they could not prepare for it.

Were your tests photocopied, or did learners see your test items or test tasks another way, such as on the blackboard?

I uploaded the instructions on how to complete the final project two months before the deadline using the university's blackboard system. In Korea, it is called E-Class, and it is the site, within the university's website, used to interact with students and share the materials/announcements/etc. with them. All the students had access to the instructions at the same time and all of

them could access them as many times as needed and even download them to their own devices if they wanted to.

How did you deal with learners who missed the test, or who were late for the test?

They knew this was their final project/exam and that, if they missed it, they would get a 0 in the presentation part. As this was the part of the evaluation that counted the most for their final grade, so nobody missed it or was late.

How did you prepare learners to take the test? Were there any test items or test procedures that were new to them?

Everything was new for them, as I am the only professor in my university who has replaced the final interview with final projects. To make sure everybody understood the main idea of the final project, I used part of one of the lessons to go over the instructions with them.

In the instructions themselves, I included several examples of videos they could watch to get ideas on how to develop this final project. And, in class, I gave an example about how a presentation, including the voice-over part of the project, should look and sound like.

Did you pilot your test? Do a trial run?

I did not, although this is an aspect I will take into consideration when preparing future final projects like the one I implemented, and am sharing in this book chapter.

Did you write any of the test in the learners' first language? Why?

No, because my Korean level is not advanced enough to write anything like what I had in the instructions. Moreover, their level of Spanish was advanced enough for them to understand the instructions, although I used some English words when I was describing the task in class for them, just to make sure everybody was following me.

Was your test administered on a computer?

No. While students used computers to prepare their presentation (e.g., editing the videos, etc.) and deliver it in front of the class, I don't think computers alone are a good way of conducting a type of test that requires learners to interact with the audience and answer questions. I think it could be doable in circumstances in which the students cannot be present in-class. However, I still favour the option of doing presentations in person.

For classroom tests: How did you accomplish scoring learners' tests?

Consider: Did you score learners' tests twice for accuracy?

Did you hide learners' names as you scored?

Did you go back and change your marks on previously scored tests in response to problems you found while scoring tests later in the process?

As I mentioned before, I knew I had to evaluate everything by myself, so I decided to evaluate the final projects myself on two different occasions.

On the first occasion, I gave them a score shortly after they had finished their presentation and I had been taking notes on their performance. Once I was back at home, I evaluated everybody's performance a second time, this time randomizing the order in which I reviewed the presentations/videos and I looked at my own notes. Normally, both evaluations were quite similar, although in some cases I realized I was being too harsh with some students (taking off points for mistakes they were making that were beyond the expectations of the course). Thus, I think this double process of evaluating them allowed me to be equally fair with everybody.

It was impossible to hide the identity of the students' whose work I was evaluating at each point, as I could see them in the videos and hear their voices in the recording I did of their presentations. However, as I mentioned, I tried to counterbalance the other in which I evaluated them to make sure no personal reasons (e.g., being tired) affected negatively any of the students.

Did you put learners' responses to items into a spreadsheet for further analysis? Did that process help you catch scoring accuracy problems? or problems with bias?

I did. I created a spreadsheet that was automatically calculating their final grade (considering the different percentages that each part of the project was worth it, such as 10% for their classmates' feedback). In this spreadsheet I included the grade for each one of the portions of the final project, as well as my notes on each specific aspect (e.g., my notes on their individual and group participation). I used these notes and grades to complete the form that I sent to each individual student explaining them their final grade.

Personally, I thought that this process made it easier for me to make sure I was being equally fair with everybody. This is how I realized in some cases I had been a little bit too harsh with some students.

Did you ask the students themselves to score their own test? Or a classmate's test?

Actually, I did. As it can be seen in the test I attached to this questionnaire, I asked all the students to individually provide feedback to the other groups

using the students' scoring rubric in Spanish (Table 4-31). I wanted to make sure students were engaged and paying attention to other students' participation and to judge these students' efforts. I did not ask them to evaluate their classmates' grammar/vocabulary (that was my job), but the effort put into their final project.

For performance tests: How did you accomplish scoring learners' performances?

As mentioned before, I knew I had to evaluate everything by myself, so I decided to evaluate the final projects myself on two different occasions. Normally, both evaluations were quite similar, although in some cases I realized I was being too harsh with some students (taking off points for mistakes they were making that were beyond the expectations of the course).

Do you think your test was reliable? What did you do to check?

This is a hard question to answer. I think my test was reliable, as the final grades obtained correlated with other measures of students' proficiency I collected during the semester (weekly homework assignments and midterm scores). However, apart from this correlation, I do not have any other "proof" of its reliability, other than my own judgement.

Reporting scores

Did you report the scores to learners? If so, how did you report test scores to learners?

I used this test to measure learners' progress and their speaking skills, and each student received the final project evaluation form (Table 4-30), including information regarding his/her individual progress, his/her participation in the group, the feedback students gave to the group regarding how original the project was, and how well they had prepared to complete this assignment.

What was your goal in reporting the test scores to learners?

My goal was to make sure students understood their grade and that they could use my comments to help them further improve their speaking skills. I gave them additional feedback on general aspects, such as "you need to rely a little bit less on your notes when presenting" and on specific aspects of their grammar/vocabulary.

Did you teach learners how to interpret their test scores?

South Korea follows a relative grading, which is a government imposed grading curve. While it depends on schools, this grading curve demands that at least 30 percent of students are destined to receive a grade of C+ or lower. So, even when students deserve to get a higher level, the grading system at the university does not allow professors to give more than a certain percentage of As, of Bs, and of Cs.

The class in which I used this test is supposed to start with students with an A2 level and prepare them to take (and pass) the B1 level (as, for example, in the DELE exams). I followed the criteria examiners from the Instituto Cervantes use when evaluating the candidates taking the DELE B1 exam (Instituto Cervantes, 2019), always taking into account the grading curve I am supposed to use at all times.

How did you use learners' scores from this test?

Consider: Were the scores for your use only?

Did learners' test scores have any positive or negative consequences for you, in terms of your institution?

The learners' scores from this test were just used by me, and there were no positive or negative consequences for me. Nobody else has had access to my scores (or the results of the test themselves) other than me.

What was the role of the test score in determining learners' grades?

Consider: How much weight did you give your test? How did you decide?

Were other measures used to decide learners' grades, besides your test?

What was the relationship of the other measures to your test?

Did your test capture some knowledge, skill, or ability the other measures did not capture?

This was the final project for the class (replacing the traditional final exam), worth 40% of students' final grade. The rest of the scoring criteria was divided between weekly assignments (20%), midterm oral exam (30%), and participation (10%).

From the beginning, I knew this final project should be the aspect to receive more credit, because it would require quite a bit of preparation/work from the students, and I wanted to compensate that effort. Moreover, it was going to be the piece of evidence that would provide me with a clear account on students' real speaking improvement, as they would not be able to prepare everything beforehand and they would need to produce free speech.

Reporting scores

How did you report scores to learners? Was timeliness of concern to you?

am sure there is plenty of room for improvement. However, I think that it achieved the purpose of making them use the language to express their own thoughts rather than just repeat sentences they had previously memorized from their grammar books.

I am not sure whether I can claim that they had changed their studying habits based on the test results. However, students came to me at the end of the semester and told me in person how much they liked my class (including this final project), because they were able to really use the language for the first time in their studies.

Did you spend time going over the test in class?

No. The final project was presented during the last lecture of the semester, so we did not have any more time together to go over the test results in subsequent classes. However, it would be ideal to have some extra time at the end of the semester to give feedback on how to improve their presentations/final projects. This is something I will try to incorporate in future classes.

Did learners ask you about the test itself (not the test scores) outside of class? If so, what did they want to talk to you about?

They mostly wanted to make sure they had understood the instructions or whether they could add some extra examples/details that, even not explicitly stated in the instructions, that could help them better prepare for their final projects.

Did learners' test scores change your teaching?

Consider: Did you change your teaching for future courses based on test results?

Not for the students who completed this final project, unfortunately. However, I think the results of this experience have changed how I approach the teaching of my conversation classes since then. After this “experiment” last Spring semester, I have included many more group activities in my classroom, most of which require students to solve some sort of problem in groups.

If you could turn back time, what would you change about your test? What would you change about your test administration?

Consider: Did learners give you feedback on the test? Did they think the test was fair, or helpful?

I think I would give myself more time to go over the results of the test with the students and give feedback that could serve not only to the

individual, but also to the whole class. I used the results of the test to give individual feedback to students.

However, I feel I could have given some more general feedback to the classroom, which could have helped students better prepare for future presentations. For example, I think I would do two projects, some sort of midterm project, which could serve as a baseline to determine students' "starting point" and the final project. Then, students could make sure to put into practice the feedback I would give them after the midterm when preparing the final project.

Section Four: Evaluating and Reviewing your Answers

To what extent do you think you've described recurrent patterns in your work with tests?

I have talked all the time about the importance of the final report I submitted with the feedback and the final scores, and the importance of using this feedback, hopefully, in a more general way, so that students can implement the feedback in future classes/presentations. I think this is something recurrent in my work with tests as it is very important for me to make sure students understand their scores and have some sort of information for them to know how to improve in future classes/in their Spanish.

I have also emphasized a couple of the times in my responses to the questionnaire the importance of the training I received in the Fundamentals of Project-Based Language Learning Online Institute, offered by the National Foreign Language Resource Center (NFLRC) at the University of Hawaii at Manoa. While I modified the examples we discussed in the institute to match my students' needs, it is evident to me that learning more about new ways of testing students was critical for me to change the way in which I approach my teaching. The training is also something recurrent in my work, as I always try to keep up-to-date with the latest pedagogical approaches.

Finally, I think have also emphasized several times how important it was for me to create a test that would help me target free production. My students were used to conversation classes in which they had to complete a writing test as their final exam. However, this was something that needed to be changed. While not perfectly, I am sure, this test is the beginning to try to set up a curriculum that emphasizes the specific skills that the classroom is meant to teach. Understanding that a test should really target the language skill practiced in the classroom is another aspect that is recurrent in my work.

To what extent is your test here an innovation, or something new, for you?

I am honestly proud of this final project, and this is the reason why I decided to submit it for publication as a book chapter. While I am completely aware of the fact that it could be improved in many ways, this was the starting point for me to change my teaching/evaluation habits.

I used to follow traditional ways to create and administer tests (e.g., oral interviews or written reports). However, I always felt that I was not fully capturing the real level of proficiency of my students, and that I was limiting their possibilities to freely express themselves. I think this type of project allows the teachers to evaluate the real level of proficiency of the students, while the students are doing something meaningful for them. Since then, I am trying to implement these ideas in all my courses.